DOCUMENT RESUME

ED 364 574 TM 020 794

AUTHOR Linacre, John Michael; Wright, Benjamin D.
TITLE Constructing Linear Measures from Counts of Qualitative Observations.
PUB DATE 93
NOTE 9p.; Paper presented at the International Conference on Bibliometrics, Informetrics and Scientometrics (4th, Berlin, Germany, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computation; Data Collection; *Goodness of Fit; Item Response Theory; Measurement Techniques; Models; *Qualitative Research; Quality Control; *Rating Scales; Research Methodology; *Test Construction
IDENTIFIERS Empirical Research; *Linear Relationships; Ordinal Position; *Rasch Model

ABSTRACT

Disorder and linearity are essential to science, but empirically derived laws indicate that counts do reflect underlying structure and can support inference. Starting from the premise that a linear structure underlies data, G. Rasch deduced that the necessary and sufficient mechanism to convert from counts to linear measures is the logistic ogive. Because linearity, and not mere numerosity, is what nearly all widely used statistical procedures assume of their data, Rasch theory enables the construction of linear measures from ordinal counts. Three variants of the Rasch model are presented for the following items: (1) dichotomies; (2) Poisson counts; and (3) rating scales. The need for data to cooperate in the construction of measures motivates the assessment and selection of useful data by means of quality control fit statistics. Rasch measurement is successfully used in many fields where it aids research by building a firm and level foundation from necessarily uncertain and uneven counts. (Contains 16 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JOHN M. LINACRE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the
Fourth International Conference on
Bibliometrics, Informetrics and Scientometrics
Berlin, Germany, 1993.

ED 364 574

Constructing linear measures from counts of qualitative observations

John Michael Linacre
Benjamin D. Wright
MESA Psychometric Laboratory, University of Chicago
5835 S. Kimbark Ave, Chicago IL 60637, USA
(312) 288-5650, FAX (312) 702-0248, E-mail: MESA@uchicago.edu

Abstract

Disorder and linearity are essential to science. Empirically derived laws indicate that counts do reflect underlying structure and can support inference. Starting from the premise that a linear structure underlies data, Rasch deduced that the necessary and sufficient mechanism to convert from counts to linear measures is the logistic ogive. Three variants of the Rasch model are presented: for dichotomies, Poisson counts and rating scales. The need for data to cooperate in the construction of measures motivates the assessment and selection of useful data by means of quality control fit statistics. Rasch measurement is successfully employed in many fields.

Replication and Disorder

The advance of science depends on replication. It was not when Roentgen[1] made the accidental discovery of a "peculiar black line" that science advanced, for "many such researchers had found their photographic plates in the lab unaccountably spoiled." It was only after Roentgen "put the new effect through its paces" and took the time to "check the facts," that X-rays became a scientific advance.

Roentgen replicated the original phenomenon. Replication combines duplication with alteration. Certain conditions are maintained. Certain conditions are different. The identical conditions maintain order. The changed conditions introduce disorder. The more that the essential nature of the original phenomenon, its "information", continues in the face of increasing disorder, the more robust is that information. "On the side of pure speculative theory I suggest that information is measured ... by the order it produces out of disorder. But order of what? The answer seems to be that each piece of information has value insofar as it relates to the order of other information, and that what we see in mapping [scientific papers onto a knowledge plane] is this basic order."[2]

TM020794

Here are two key ideas: 1) disorder is necessary to measure information; 2) useful measurement enables a location to be established. Though location on a plane in two-dimensions can be more instructive than location on a line in a single dimension, the common requirement is equal-interval, linear measurement.

Empirically-derived Measurement

The struggle to wrest useful order from the hodgepodge of data that inundates the scientist devolves ultimately into a quest for linear measurement. If we cannot measure a phenomenon in equal-interval units, there is grave doubt that we understand it in any scientific sense. "Length" and "Love" are cases in point. "Indeed, it is a character of all the higher laws of Nature to assume the form of precise quantitative statement."[3]

Bradford's Law of Scattering demonstrates equal-interval measurement, linearity, over the central part of its range. Bradford obtained this linearity by arbitrary manipulation of the empirical data. Garfield observes that "Bradford's, Zipf's, Lotka's. and Pareto's laws were independently formulated to explain disparate phenomena. It seems more than coincidence that they should closely resemble each other. One wonders if they are not all governed by a single underlying principle."[4] As Bookstein[5] contends, the existence of Bradford's law proves that a more general form of regularity must be present in the data, of which Bradford's law is a special case, or else Bradford could not have discovered his law.

Theory-derived Measurement

The empirical approach is to collect whatever there is in the way of data and then to hunt for some description that summarizes those data in a simple form. The theory-based approach is to develop the criteria that must be satisfied in order for linear measures to be constructed, and then to learn how to construct data that satisfy those criteria.

The theory-based approach speeds the scientific process because it leads to the elimination of superficially attractive, but fundamentally unproductive, statistical manipulations. It also permits investigators to locate coherent subsets of data within otherwise incoherent sets.

Counting: a qualitative decision

Linear measurement begins with counting. Linear measures, e.g., grams, are ideals only imperfectly realized in practice. Counts are empirical absolutes. The observer can count ten apples on a tree, or ten papers by an author. Clearly, not all apples are equally large and crisp, not all papers are of equal value (p.40).[6] The first question for the most basic quantitivity of counting is thus qualitative. Are the objects counted sufficiently alike so as to be considered interchangeable in the counting process? Stricter rules lead to more discriminating counting processes. Is a rotten apple counted as an apple? Is a corrected version of an earlier paper counted as another paper?

Air lines have discovered that there is a convenient, approximate, but useful and accurate enough conversion from passenger count to passenger weight. Passenger count is an ordinal, discrete integer summarizing passengers with a wide range of characteristics. Passenger weight is an interval-scale amount that idealizes one aspect of the passengers. Once the qualitative decision has been made as to who to count, the particular details of each passenger become irrelevant.

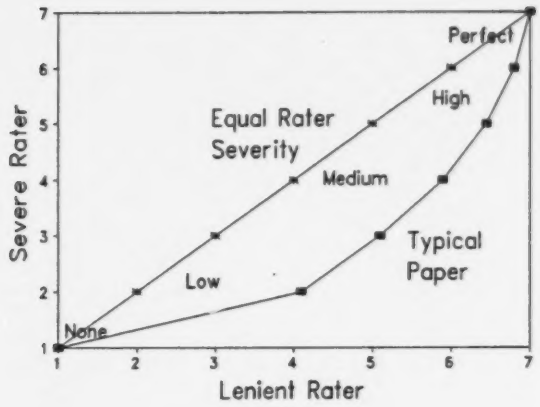
The essential element in making scientific inferences from counts is that the same mechanism converts counts to linear measures regardless of the counting rule employed. Thus counts of green apples, red apples, big apples or ripe apples are all converted to linear weights by the same mathematical process. Only the conversion factors differ.

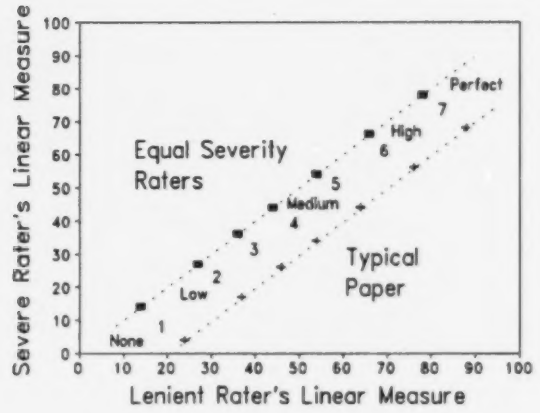
Linear measurement: an idealization

It is evident that a single dichotomous observation of a variable, such as the presence or absence of a particular author among the list of names heading a paper, is not a linear measure of the proclivity to authorship of that author. Nevertheless, an accumulation of such dichotomous observations into a count summarizes all the available information that exists about that author's proclivity.

Table 1. COUNTS ARE NOT LINEAR MEASURES

CHARACTERISTIC	COUNTS	LINEAR MEASURES
MEANING	Ordinal rank on hoped-for variable	Linear position on explicitly defined variable
STATUS	mistaken for "truth"	known as estimates
ADDITIVITY	non-linear, bent	linear, straight
CONTINUITY	discrete, lumpy	continuous, smooth
PRECISION	unknown	quantified
LINEAR ANALYSIS	unsuited to usual statistics	ideal for usual statistics
CONCEPTUALIZATION	concrete	abstract
OCCURRENCE	accidental	deliberate
CONSTRUCTION	irreproducible	reproducible
SCOPE	local	general
CONTEXT	situation-bound	situation-free





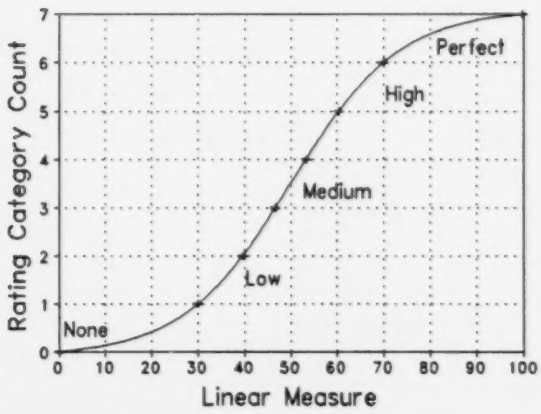


Figure 1. Non-linearity of counts. A, non-linear relationship between rating levels of raters with different severities. B, linear relationship between linearized measures corresponding to rating levels of different raters. C, relationship of counts to linear measures.

A count is often treated as or assumed to be a linear measure. But the mere fact that counts exhibit numerosity and can be written as linear numbers does not give them linear properties. Table 1 contrasts the properties encountered in counts with those required of linear measures. Counts often approximate linearity over a central range closely enough for researchers to ignore their deficiencies. Nevertheless this lack of linearity has usually proved fatal to meaningful theory development.

Consider journal articles rated on a number of criteria by two experts. One expert applies the criteria leniently and loosely, the other severely and strictly. The ratings for each article are accumulated for each rater and become counts. The non-linearity of the counts is shown in the Lorenz-like curve in Figure 1A where equivalent counts by the lenient and severe raters are plotted. These counts may approximate linearity over their central ranges, but provide only weak support for inference from the counts of the lenient rater to those of the severe rater. Counts by both raters must be linearized into measures and then equated, as shown in Figure 1B, before inferences based on the lenient rater's counts can be applied with confidence to the severe rater's counts.

Figure 1C depicts the general ogival relationship between measures and counts. The step from counts to linear measures is one that empiricists find by accident. The logistic ogive or "autocatalytic curve" is widely used to describe physical processes. "We are lead to suggest a second basic low of the analysis of science: all the apparently exponential laws of growth must ultimately be logistic." (p. 30)[6] Nevertheless, it is not some physical mechanism that connects the counts with the measures, it is a mathematical property that only the logistic transformation possesses.

Rasch Measurement Models

Rasch[7] constructed a measurement theory that contains the necessary and sufficient criteria for the conversion of counts of qualitatively similar objects into quantitatively linear measures. This theory capitalizes on the disorder present in observations due to the stochastic element that exists in all empirical counting operations. Rasch started with the premise that the linear measure of an object, e.g., an author's proclivity to publish in a special field, must be independent of the measurement agent, e.g, any particular journal, and *vice-versa*. Each object has a single linear measure. Each agent a single linear measure, its calibration, in that same frame of reference. When measures and calibrations are hypothesized to interact to produce stochastic data, a necessary and sufficient measurement model can be derived for any type of count.

Table 2. EXAMPLES OF RASCH MEASUREMENT MODELS

Type of Data	Observed Categories	Ordinal Interpretation	Measurement Model for $\text{Log}[P_{nij}/P_{nij-1}] = ?$
Dichotomies	no = 0 yes = 1	less more	$B_n - D_i$ $j=1$
Poisson Counts	0 ∞	0 ∞	$B_n - D_i - F_i * \log[j]$ $j=1,\infty$
Likert Ratings	1 2 m	least most	$B_n - D_i - F_j$ $j=2,m$

The essential transformation is always the logistic, log-odds transformation. Table 2 lists measurement models for three situations.[8] The basic model for dichotomous, true/false, present/absent 1/0 observations is

$$\log\left(\frac{P_{ni}}{1 - P_{ni}}\right) = B_n - D_i$$

Object n is parameterized to have linear measure B_n. Agent i is parameterized to have linear calibration D_i. These interact stochastically to produce a datum with probability P_{ni} of being observed to be 1.

When the data are Poisson counts of independent events, F_i is the conversion factor from the logarithm of the count to the general linear scale for the particular agent of investigation, i. This formulation supports most empirical laws based on logarithms of counts.

Likert ratings are observations on a rating scale or some other ordered series of categories. Observations are ordinal because it is specified that the higher the particular observation, the more of the variable is manifested. This model underlies Figure 1. The inevitably uneven spacing of the categories is parameterized by F_j, the extra "difficulty" overcome in order to be observed in category j, relative to category j-1. The count-to-measure transformation again has the appearance of a logistic ogive, but its precise form depends on the behavior of the rating scale categories.

Measurement and Quality Control

A major difference between research in the social sciences and that in other sciences is quality control. Physicists take many observations, rejecting those that seem dubious, and keeping those of the high quality.[8] In bibliometrics, the need to verify the quality of data is well established: "As we have repeatedly done in the past, may we respectfully caution against the serious

implication ... that quantitative data [i.e., counts] can be used without considered (not rote) qualitative judgments."[10] Most social scientists, on the other hand, attempt to include in their analysis every observed datum. They perform global fit tests in which the data is treated as sacrosanct, but it is the theoretical model that is accepted or rejected *in toto*. Rasch measurement rejects the "global fit" approach in favor of detailed quality control.

For the construction of linear measures, quality control through detailed examination of the fit of the data to the measurement model is vital. Here it is essential to apply Maier's tongue-in-cheek law: "If facts do not conform to theory, they must be disposed of."[11] For linear measures to be obtained, the counts must fit a Rasch model. For each analysis, the decisive question is not "Does the theoretical model fit the empirical data?", but "Do the data fit the model usefully enough to support linear measurement?"

Rasch models provide quality-control fit statistics at all levels of analysis. Individual data points can be assessed for their likelihood. Objects and agents, e.g., authors and journals, can be scrutinized for their adherence to the general behavior of objects and agents. Groups of objects and agents can be investigated for distinguishing characteristics.

The nature of the lack of fit motivates the remedial action. Idiosyncratic, non-cooperative data points are isolated for diagnostic investigation. Are they data entry errors? Are they unusual circumstances? Data-to-model misfit prompts deeper questions: do the data manifest quantities of one predominant underlying variable, or are they a heterogeneous collection of observations? Can homogeneous subsets be identified and measures for these constructed? Irremedial lack of fit signifies that those counts must be rejected as not supporting linear measurement.

The Successful history of Rasch Measurement

Rasch[7] developed and expanded his measurement methodology between 1951 and 1959. Early work included linear measurement of oral misreadings, readings speed and math ability. This lead to major applications of Rasch measurement in educational and psychological testing. Though initially applied to dichotomous multiple-choice tests, the model is now employed, in its rating scale form, for performance assessment. The robustness of the model against missing data, has made it ideal for use with computer-adaptive tests.

The ready availability of software, e.g., BIGSTEPS[12] and *Facets*[13], has promoted the application of Rasch measurement to many fields including quality of paint finishes[14], performance of wine-tasters, measurement of evolutionary development in fossils, and medical patient evaluation.[15,16]

Conclusion

Rasch theory enables the construction of linear measures from ordinal counts. Linearity, not mere numerosity, is what nearly all widely used statistical procedures assume of their data. The computation of meaningful means, standard deviations, and least-squares estimates requires linearity in the underlying data.

Once linear measures have been constructed, regression analysis, ANOVA and even simple plots have a firm foundation. Without linear data, it is not known whether the regression coefficients are explaining the underlying phenomenon or merely the non-linearity of the counts. Without linearity, it cannot be determined whether "ceiling" and "floor" effects are substantive findings or merely artifacts of the way the counts were made. Rasch methodology aids research by building a firm and level foundation from necessarily uncertain and uneven counts.

References

1. D. De Solla Price, Science since Babylon, Yale University Press, New Haven, 1961, p. 77-78.
2. D. de Solla Price, Foreword to E. Garfield, Essays of an Information Scientist, Vol 3, 1977-1978, ISI Press, Philadelphia, 1980, p. vii.
3. J. F. W. Herschel, Preliminary discourse on the study of natural philosophy, Longman, Brown, Green & Longmans, London, 1851, Art. 116.
4. E. Garfield, Bradford's Law and Related Statistical Patterns, Essays of an Information Scientist, Vol 4, 1979-1980, ISI Press, Philadelphia, 1981, p. 476-483.
5. A. Bookstein, Informetric Distributions, Part 1: Unified Overview, Journal of the American Society for Information Science, 41 (1990), 5, 368-375.
6. D. De Solla Price, Little Science, Big Science, Columbia University Press, New York, 1963
7. G. Rasch, Probabilistic models for some intelligence and attainment tests, University of Chicago Press, 1980.
8. B.D. Wright, G. N. Masters, Rating Scale Analysis, MESA Press, Chicago, 1982.
9. N. E. Dorsey, The Velocity of Light, Transactions of the American Philosophical Society, 34 (1944), 1-110.
10. E. Garfield, Citation measures used as an objective estimate of creativity, Essays of an Information Scientist, Vol 1, 1962-1973, ISI Press, Philadelphia, 1977, p. 120-121.
11. E. Garfield, Humor in scientific journals and Journals of Scientific Humor, Essays of an Information Scientist, Vol 4, 1974-1976, ISI Press, Philadelphia, 1977, p. 664-670.
12. B. D. Wright, J. M. Linacre, BIGSTEPS Rasch analysis computer program, MESA Press, Chicago, 1991.
13. J. M. Linacre, *Facets* many-facet Rasch analysis computer program, MESA Press, Chicago, 1987.
14. T. K. Rehfeldt, Judgement of stain resistance, Journal of Coatings Technology, 62 (1990), 790, 53-58.
15. S. M. Haley, L. H. Ludlow, Applicability of the Hierarchical Scales of the Tufts Assessment of Motor-Performance for School-Aged Children and Adults with Disabilities, Physical Therapy, 72 (1992), 3, 191-202.
16. C. V. Granger, B. B. Hamilton, J. M. Linacre, A. W. Heinemann, B. D. Wright, Performance profiles of the Functional Independence Measure, American Journal of Physical Medicine and Rehabilitation, 72 (1993), 2, 84-89.